American Journal of Epidemiology
Vol. 164, No. 3
DOI: 10.1093/aje/kwj183
Advance Access publication May 17, 2006

Invited Commentary

Invited Commentary: On the Road to Improved Exposure Assessment using Geographic Information Systems

Mary H. Ward[1] and Daniel Wartenberg[2]

[1] Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Bethesda, MD.
[2] Division of Environmental Epidemiology, Department of Environmental and Occupational Medicine, University of Medicine and Dentistry of New Jersey–Robert Wood Johnson Medical School, Piscataway, NJ.

Received for publication October 4, 2005; accepted for publication March 20, 2006.

Abbreviation: GIS, geographic information systems.

The use of geographic information is becoming more commonplace in epidemiologic research. Epidemiologists use geographic information systems (GIS) for both designing studies and analyzing data. For example, GIS are used in disease mapping (1–5), identifying potential populations for study (6), conducting small-area analyses of disease rates (7), and undertaking disease cluster and surveillance activities (8–12). Exposure assessors use GIS for developing high-resolution estimates of environmental exposures through data analysis and modeling, for example, for air pollutants (13), pesticides (14–16), and water pollutants (17). Such applications require the consideration of issues that span the disciplines of geospatial science, environmental science, and epidemiology (18). Geospatial issues include the spatial scale and resolution (positional accuracy) of the exposure and health outcome data. Relevant environmental science issues include the fate and transport of specific contaminants in the environment and the validity of the geographic model used to estimate exposure. Finally, issues that must be considered in the epidemiologic study design and analysis include the evaluation of potential confounders and concurrent exposures to multiple risk factors, the etiologic relevance of the exposure levels and exposure timing, and consideration of the disease latency. The estimation of the errors in the exposure metrics and the effects of exposure misclassification or bias on risk estimates are also important considerations in the interpretation of study findings (19).

In this issue of the *Journal*, Yu et al. (20) present results from an innovative study in which they investigate the association of leukemia in children and young adults with residential exposure to petrochemicals. The study is noteworthy, in part, because of the way it uses GIS tools to estimate individual-level exposures and to assess exposure-disease relations. This is the latest in a series of investigations by these authors in which they evaluate the occurrence of cancer and other adverse health outcomes in populations living near petrochemical facilities in Taiwan (21–24). This study is an improvement on their previous work, because the investigators have geocoded both residence and exposure-source locations, modeled the transport of the contaminants from source to residence, and used an individual-based (rather than an aggregate or ecologic) design for the assessment of the possible association between exposure and disease. Thus, their study design addresses many issues implicit in using GIS in environmental epidemiology, and we discuss a few of these. Specifically, we focus on the positional accuracy of the exposure data, the validity of the exposure model, and the use of individual rather than aggregate exposure data for epidemiologic analysis.

To estimate residential exposures, Yu et al. (20) first used GIS to geocode the study population's residential addresses over the time period relevant to the disease. In any study, excluding participants whose addresses cannot be geocoded can introduce bias (25, 26). Therefore, it is important that the assumptions and methods in the geocoding process be documented so that errors and the potential for differential exposure misclassification can be evaluated. Toward this end, Yu et al. presented the success of the

Correspondence to Dr. Mary H. Ward, Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Boulevard, EPS 8104, Bethesda, MD 20892 (e-mail: wardm@exchange.nih.gov).

geocoding effort among cases and controls, and they considered the potential for differential exposure misclassification among responding and nonresponding controls. Such details are often ignored in the presentation of epidemiologic studies using GIS; however, this type of information is essential for a critical evaluation of the study results and should be reported routinely, similar to the way in which participants' response rates are reported in most epidemiologic studies.

In addition to reporting the completeness of the geocoding, it is important to identify the base maps used for geocoding the participants' residences and the exposures' sources (e.g., point sources, roadway coordinates for traffic, well locations). Differences from tens to thousands of feet may exist between geographic coordinates for the same site on similar maps from different sources; therefore, using discordant base maps could result in substantial exposure misclassification. Although it is preferable to use the same base maps and geocoding software for geocoding all residences and exposure sources, one can assess consistency across maps or methods by conducting a sensitivity analysis by geocoding a subset of locations using all the geocoding methods and base maps and comparing the positional results. Differences may be trivial or may be substantial. To evaluate the accuracy of the geocoding process, aerial photographs or other ground truth information can be used. Note that the positional accuracy of geocoding can be greater for urban addresses compared with rural addresses (27–30); therefore, assessing the accuracy of the geocoding process may be particularly important when the residences are located across a study area with a wide range of population densities.

To estimate exposure, Yu et al. (20) calculated residential distances to the multiple petrochemical facilities over the lifetime of the study participants. They weighted the distance to the facilities by the prevailing wind direction, rather than simply using an unweighted distance to the single, closest facility (i.e., a simple residential proximity metric). Many studies have used residential proximity to pollution sources as a surrogate for exposure (for a review, refer to reference 18). Assumptions inherent in that approach include that exposure decreases as a linear or inverse square function of the distance from the source, that wind direction patterns are symmetric around the source, and that the terrain is relatively flat (31). Since these assumptions are violated in many if not most study areas, a simple proximity metric has many limitations. The exposure metric used by the authors is far better, accounting for multiple emission sources, the monthly prevailing wind direction, and assuming exposure decreases in relation to the inverse of the distance from the facilities up to 3 km away. Although a substantial improvement over proximity alone, this distance model, too, has limitations. In particular, the heights of the petrochemical facility's smokestacks were not modeled, even though variation in this parameter can affect the transport of emissions. A Gaussian-plume straight-line model would likely further improve the exposure estimates. Such models have been used for predicting concentrations downwind from smokestacks in relatively flat terrain and are easily incorporated into GIS (31). A further refinement might consider the aerodynamic and chemical properties of emissions, as well as wind speed. Not all airborne substances move and/or disperse equally, and some transform chemically over time. For example, particulates settle differentially as a function of size, shape, and density, and this varies further with changes in wind speed. Consideration of these issues also could improve the accuracy of the exposure estimates.

Validating an exposure model also is an important part of exposure assessment. Including monitoring data in the development and calibration of the geographic model can result in improved model emissions estimates, as has been recommended by the National Research Council (32). This can be accomplished, for example, by using ambient air-monitoring data to benchmark the model estimates. Previous studies have described environmental monitoring data for polycyclic aromatic hydrocarbons and volatile organics in the study area (20). Comparing these data with the authors' exposure estimates would provide one measure of accuracy and consistency over space and time.

In addition to the transport of chemicals in the environment, the likely route of exposure should be considered. Other sources of exposure to contaminants, including traffic, additional stationary sources, and indoor exposures, also could be included in estimates of cumulative exposures, or at least their relative contributions should be assessed.

In an effort to reduce the exposure misclassification, error, and bias that are common in ecologic designs, Yu et al. (20) estimated exposure to airborne emissions from petrochemical facilities in southern Taiwan at each subject's current and past residences. Most epidemiologic studies that have used GIS for estimating environmental exposures (18, 31) have used ecologic study designs, despite their limitations (33), often because ambient environmental quality data and health outcome data were available only at the regional level. In this way, the study also represents a major improvement over many previous GIS-based epidemiologic studies. Because an ecologic analysis could be conducted using the data already acquired for their individual study, it would have been interesting to compare aggregate and individual estimated effect measures and to quantify the improvement in the precision of the estimates.

Further refinement of exposure estimation would take into account the activity patterns of the study participants, including where they went to school and work. Personal exposure levels are influenced by all these factors and, as a result, personal exposure measures can differ greatly from outdoor air concentrations (34, 35). A critical step in improving exposure assessments is the incorporation of global positioning system (GPS) technology for monitoring activity patterns into epidemiologic study designs (18, 36). An example is a method developed by Gulliver and Briggs (37) to estimate traffic-related air pollution using time-space modeling of journey-time exposures. In addition, use of personal monitoring devices for selected subpopulations, as have been used in air pollution (38) and magnetic field (39) exposure studies, can help to link ambient concentrations to personal exposure levels. Such approaches will bring us closer to the ultimate goal of accurately estimating personal exposure. Personal measurement data are the ideal exposure metrics for determining personal exposure. However, for many environmental contaminants, personal monitoring is not feasible or is not representative of past exposures. Therefore, the continued evaluation of environmental exposures and chronic disease risk will rely on geographic models, GIS, and global positioning system technology.

Progress toward reaching the full potential of GIS applications in environmental epidemiology will likely take place in stages as more environmental monitoring data become available, and as more sophisticated models and modeling approaches are developed. The increasing availability of high-resolution, spatially registered environmental monitoring data and the accessibility of GIS software and spatial analysis methods will facilitate further refinement of exposure models and improved accuracy in estimated exposures. The extent to which the full potential of GIS is utilized in both the study design and the exposure assessment will determine the success of future environmental epidemiology studies in identifying and characterizing disease etiologies. We believe that, by changing and improving the exposure assessment methodology as new technology and data are made available, epidemiologists will be able to improve the accuracy, precision, and sensitivity of their investigations of the etiology of environmentally caused disease.

ACKNOWLEDGMENTS

Conflict of interest: none declared.

REFERENCES

1. Mason TJ. Cancer mortality in U.S. counties with plastics and related industries. Environ Health Perspect 1975;11:79–84.
2. Pickle LW, Mungiole M, Jones GK, et al. Atlas of United States mortality. Hyattsville, MD: National Center for Health Statistics, 1996.
3. Aylin P, Maheswaran R, Wakefield J, et al. A national facility for small area disease mapping and rapid initial assessment of apparent disease clusters around a point source: the UK Small Area Health Statistics Unit. J Public Health Med 1999; 21:289–98.
4. Devesa S, Grauman DJ, Blot WJ, et al. Cancer surveillance series: changing geographic patterns of lung cancer mortality in the United States, 1950 through 1994. J Natl Cancer Inst 1999;91:1040–50.
5. Lawson AB, Biggeri A, Boehning D, et al, eds. Disease mapping and risk assessment for public health. New York, NY: Wiley, 1999.
6. Wartenberg D, Greenberg M, Lathrop R. Identification and characterization of populations living near high-voltage transmission lines: a pilot study. Environ Health Perspect 1993;101: 626–32.
7. Elliot P, Wartenberg D. Spatial epidemiology: current approaches and future challenges. Environ Health Perspect 2004;112:998–1006.
8. Hardy RJ, Schroder GD, Cooper SP, et al. A surveillance system for assessing health effects from hazardous exposures. Am J Epidemiol 1990;132(suppl):S32–42.
9. Wartenberg D. Should we boost or bust cluster investigations? (Editorial). Epidemiology 1995;6:575–6.
10. Rushton G, Lolonis P. Exploratory spatial analysis of birth defect rates in an urban population. Stat Med 1996;15: 717–26.
11. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. J R Stat Soc (A) 2001;164: 61–72.
12. Rogerson P. Monitoring point patterns for the development of space-time clusters. J R Stat Soc (A) 2001;164:87–96.
13. Jarup L. Health and environment information systems for exposure and disease mapping, and risk assessment. Environ Health Perspect 2004;112:995–7.
14. Ward MH, Nuckols JR, Weigel SJ, et al. Identifying populations potentially exposed to agricultural pesticides using remote sensing and a geographic information system. Environ Health Perspect 2000;108:5–12.
15. Brody JG, Vorhees DJ, Melly SJ, et al. Using GIS and historical records to reconstruct residential exposure to large-scale pesticide application. J Expo Anal Environ Epidemiol 2002;12:64–80.
16. Gunier RB, Harnly ME, Reynolds P, et al. Agricultural pesticide use in California: pesticide prioritization, use densities, and population distributions for a childhood cancer study. Environ Health Perspect 2001;109:1071–8.
17. Schwartz CH, Rudel RA, Kachajian JR, et al. Historical reconstruction of wastewater and land use impacts to groundwater used for public drinking water: exposure assessment using chemical data and GIS. J Expo Anal Environ Epidemiol 2003;13:403–16.
18. Nuckols JR, Jarup L, Ward MH. Using GIS for exposure assessment in environmental epidemiology studies. Environ Health Perspect 2004;112:1007–15.
19. Brenner H, Greenland S, Savitz DA. The effects of nondifferential confounder misclassification in ecologic studies. Epidemiology 1992;3:456–9.
20. Yu CL, Wang SF, Pan PC, et al. Residential exposure to petrochemicals and the risk of leukemia: using geographic information system tools to estimate individual-level residential exposure. Am J Epidemiol 2006;164:200–7.
21. Yang CY, Cheng MF, Chiu JF, et al. Female lung cancer and petrochemical air pollution in Taiwan. Arch Environ Health 1999;54:180–5.
22. Yang CY, Chiu HF, Chiu JF, et al. Cancer mortality and residence near petrochemical industries in Taiwan. J Toxicol Environ Health 1997;50:265–73.
23. Yang CY, Wang JD, Chan CC, et al. Respiratory and irritant health effects of a population living in a petrochemical-polluted area in Taiwan. Environ Res 1997;74:145–9.
24. Pan BJ, Hong YJ, Chang GC, et al. Excess cancer mortality among children and adolescents in residential districts polluted by petrochemical manufacturing plants in Taiwan. J Toxicol Environ Health 1994;43:117–29.
25. Hurley SE, Saunders TM, Nivas R, et al. Post office box addresses: a challenge for geographic information system-based studies. Epidemiology 2003;14:386–91.
26. Gregario DI, Cromley E, Mrozinski R, et al. Subject loss in spatial analysis of breast cancer. Health Place 1999;5: 173–7.
27. Ward MH, Nuckols JR, Giglierano J, et al. Positional accuracy of two methods of geocoding for rural and community addresses. Epidemiology 2005;16:542–7.
28. Bonner MR, Han D, Nie J, et al. Positional accuracy of geocoded addresses in epidemiologic research. Epidemiology 2003;14:408–12.
29. Cayo MR, Talbot TO. Positional error in automated geocoding of residential addresses. Int J Health Geogr 2003;2:10.

(doi:10.1186/1476-072X-2-10) (http://www.ij-healthgeographics.com/content/2/1/10).
30. Rushton G, Armstrong MP, Gitller J, et al. Geocoding in cancer research: a review. Am J Prev Med 2006;30(suppl):S16–24.
31. Beyea J, Hatch M. Geographic exposure modeling: a valuable extension of geographic information systems for use in environmental epidemiology. Environ Health Perspect 1999; 107(suppl 1):181–90.
32. National Research Council. Human exposure assessment for airborne pollutants: advances and opportunities. Washington, DC: National Academies Press, 1991.
33. Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. Annu Rev Public Health 1995;16: 61–81.
34. Ott WR. Human exposure assessment: the birth of a new science. J Expo Anal Environ Epidemiol 1995;5:449–72.
35. Payne-Sturges DC, Burke TA, Breysse P, et al. Personal exposure meets risk assessment: a comparison of measured and modeled exposures and risks in an urban community. Environ Health Perspect 2004;112:589–98.
36. Weis BK, Balshaw D, Barr JR, et al. Personalized exposure assessment: promising approaches for human environmental health research. Environ Health Perspect 2005;113:840–8.
37. Gulliver J, Briggs DJ. Time-space modeling of journey-time exposure to traffic-related air pollution using GIS. Environ Res 2005;97:10–25.
38. Brauer M, Hirtle RD, Hall AC, et al. Monitoring personal fine particle exposure with a particle counter. J Expo Anal Environ Epidemiol 1999;9:228–36.
39. McBride ML, Gallagher RP, Theriault G, et al. Power-frequency electric and magnetic fields and risk of childhood leukemia in Canada. Am J Epidemiol 1999;149:831–42.